Data Fusion in Large Arrays of Microsensors[1]

Alan S. Willsky
Room 35-433, M.I.T.
Cambridge, MA 02139
willsky@mit.edu

I. Introduction

This paper provides a brief summary of the research being performed under the ODDR&E MURI on the topic of data fusion in large arrays of microsensors, a program that involves researchers from MIT, The University of Illinois, and Princeton University.

The overall objective of this program is the investigation and development of innovative concepts and solutions to the challenges presented by the envisioned availability of very large numbers of heterogeneous sensors each of which is limited in sensing capability, power, computation, and communication. As we see it, these challenges can be grouped into three interrelated **intellectual themes (IT's)**:

IT-1: Consistent (or manageably inconsistent) fusion of networked, myopic sensors.

IT-2: Fusion and self-organization of heterogeneous sensors in unstructured and uncertain environments.

IT-3: Wireless networks, network communication and information theory.

In addition, the original description of this MURI topic listed seven **research concentration areas (RCA's)**, which represent an alternate parsing (with differing levels of granularity) of the interrelated research questions that form this important area of inquiry. The following is a list of concise statements of these RCA's:

RCA-1: Self-calibration

RCA-2&3: Fundamental limits on fusion, network information theory, tradeoffs in local vs. global processing

RCA-4: Bounds & characteristics of algorithms to ID minimum resources needed to detect, estimate, track?

RCA-5: Fusion Algorithms

RCA-6: Distributed Algorithms with guarantees on global behavior (both positive & negative!)

RCA-7: Create events/data for experiments and demos

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
00 Oct 2001	N/A	-

Title and Subtitle	Contract Number
Node Placement Guidelines for Distributed Ground Sensor Systems	Grant Number
	Program Element Number

Author(s)	Project Number
	Task Number
	Work Unit Number

Performing Organization Name(s) and Address(es)	Performing Organization Report Number
BAE Systems Advanced Systems and Technology Information and Electronic Warfare Systems P.O. Box 868 Nashua, NH 03061-0868	

Sponsoring/Monitoring Agency Name(s) and Address(es)	Sponsor/Monitor's Acronym(s)
Department of the Army, CECOM RDEC Night Vision & Electronic Sensors Directorate AMSEL-RD-NV-D 10221 Burbeck Road Ft. Belvoir, VA 22060-5806	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
See also ADM201471, Papers from the Meeting of the MSS Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors (2001) Held in Applied Physics Lab, Johns Hopkins Univ, Laurel, MD on 24-26 Oct 2001. Volume 2 (Also includes 1999 and 2000 Meetings), The original document contains color images.

Abstract

Subject Terms

Report Classification	Classification of this page
unclassified	unclassified

Classification of Abstract	Limitation of Abstract
unclassified	UU

Number of Pages
10

Section 2 of this paper contains brief descriptions of the research that has been or is currently being performed under this grant. As one would expect, individual research topics do not parse exactly into IT's or RCA's. However, we have indicated both the "center of gravity" for each topic (in terms of IT's and RCA's) as well as other elements on which each topic touches (or will touch as research proceeds in the future) and the interrelationships among these efforts.

We have established a website for this project (http://sensorweb.mit.edu), which contains information on each of the principals involved in this program, on active and completed research efforts, lists of (and links to) publications, and recent events and highlights. We refer the reader to this website for more details and up-to-date information about our activities.

II.   Research Activities

In this section we provide brief descriptions of a sampling of the research projects that are part of our MURI effort.

Topic: **Approximate Inference in GMRFs by Recursive Cavity Modeling**
By: Jason K. Johnson (MIT), Alan S. Willsky (MIT)
IT-1, RCA-5 (principal), with ties to RCA-6, 2 & 3
This research is aimed at developing near-optimal, scalable inference algorithms for random fields which admit compact description as graphical models. Estimation of Gauss-Markov random fields is the current focus of this work. The approach being developed, referred to as recursive cavity modeling (RCM), combines ideas from the multi-scale modeling theory, developed by researchers in our group before, with extensions of existing approximate inference techniques which operate by frontier propagation (FP). FP involves the construction and propagation of a frontier model which provides a faithful yet compact graphical model for the surface of one subfield for the purpose of inferring another. While FP performs inference with respect to a chain model of the field, RCM uses a tree model. The RCM approach has the advantage of decomposing the field in a more natural manner and of being ideally suited for distributed processing. Possible future extensions of this work might include the consideration of dynamic random fields as well as more general classes of static random fields.

Topic: **Stochastic Processes defined by Graphical Models**
By: Martin Wainwright (MIT), Erik Sudderth (MIT), Alan S. Willsky (MIT), Tommi Jaakkola (MIT)
IT-1, RCA-5, 6, with ties to RCA-4
Publications: [1, 2, 3, 5]
In broad overview, this research addresses stochastic processes defined by graphical models, and associated problems of modeling and estimation. The nodes of a singly-connected (i.e., tree) graph can be partially ordered in scale, which gives rise to powerful algorithms for many problems. In contrast, straightforward solution to these same problems are prohibitively complex for more general graphs of any substantial size. As a result, there is considerable interest in developing tractable techniques for graphs with loops. More specifically, we are interested in the following problems: (a) An embedded trees (ET) algorithm for exact estimation of Gaussian processes on graphs with cycles. (b) Techniques for approximate

estimation of discrete processes on graphs with cycles. Report in preparation. (c) Modeling natural image statistics with random cascades of Gaussian scale mixtures on graphical models, and their application to image processing problems.

Topic: **Modeling and Estimation of Gaussian Processes on Graphs with Cycles**
By: Erik Sudderth (MIT), Martin Wainwright (MIT), Alan S. Willsky (MIT)
IT-1, RCA-5, 6, with ties to RCA-2 & 3
Publications: [5]
Statistical models of two-dimensional fields play an important role in a variety of application areas, from oceanography to image processing. Markov random fields, perhaps the most common means of representing a two-dimensional random process, have an intuitively appealing structure, but they generally lead to computationally intensive estimation algorithms. The multiscale tree models previously introduced by the Stochastic Systems Group provide highly efficient estimation procedures. The fundamental source of the efficiency of tree-based algorithms is that there is a single unique path between any two points in the tree - in other words, there are no loops. Unfortunately, the tree structure also tends to lead to blocky artifacts in the final estimation results. This research involves the investigation of model structures that lie in between densely connected Markov random fields and singly connected multiscale trees. We have developed a novel estimation algorithm which computes the exact means and error covariances for Gaussian estimation problems on arbitrary cyclic graphs. It works by exploiting the presence of tree-like structures in more densely connected graphs. Current work focuses on obtaining a deeper understanding of this algorithm's dynamics, and on determining efficient procedures for constructing models on graphs with cycles.

Topic: **Multiresolution Modeling from Data and Partial Specifications**
By: Dewey Tucker (MIT), Alan S. Willsky (MIT)
IT-2, RCA-5
In this research, we study a recently developed class of models, called multiscale models, which is well-suited to represent a wide variety of random processes. The advantage of this type of model is in providing an extremely efficient estimation algorithm which is a generalization of the traditional Kalman filter and Rauch-Tung-Striebel smoother. Multiscale models and the associated estimation algorithm have been shown to be useful in a number of applications including image processing, remote sensing, and geophysics. In particular, we focus on the problem of constructing multiscale models from data and partial covariance information. Previous research in this area has provided an algorithm which builds a multiscale model to match the covariance of a given process of interest. However, this approach requires complete knowledge of the covariance and the capability to store it, and for problems of even moderate size, this type of complete characterization is an overwhelmingly large data storage problem. To circumvent this problem, we seek methods to construct multiscale models based solely on realizations (sample paths) of the process, with no assumed knowledge of the covariance. In addition, we examine the problem of positive definite covariance extension with the specific goal of constructing multiscale models from partial covariance information.

Topic: **Active information retrieval**
By: T. Jaakkola (MIT), H. Siegelman (Technion/MIT)
IT-2, RCA-4
We propose a new approach to effective information retrieval. The interacting agent/user is successively queried for distinctions at varying levels of abstraction and is permitted to respond with multiple selections or may choose not to respond. The information is in each

case unambiguously interpreted and incorporated by the system. The next query is chosen optimally to minimize the need for any further exchange. The system is also capable of determining whether or not the information of interest is in the (portion of the) database being consulted. In more technical terms, we make use of a stochastic substitution matrix that is derived, e.g., from local relations among the elements stored in the database. The notion of substitution permits us to define what is meant by the optimal interaction as well as construct algorithms to achieve it. All the queries are carried out under resource constraints which pertain either to the amount of information presented to the user per iteration or to the maximum/average number of iterations that are allowed. The effect of the resource constraints can be assessed prior to the interaction by using the properties of the substitution matrix as well as bounds on the successive reduction in the uncertainty about the information being sought.

Topic: **Sensor Arrays and Information Theoretic Fusion**
By: Alex Ihler (MIT), John Fisher (MIT), Alan Willsky (MIT)
IT-2, RCA-5, 6
Large arrays of sensors with overlapping regions of influence and multiple sensing modalities pose new problems in data fusion. Complex relationships between the sensor observations, including reflections and nonlinear effects, make it difficult to determine such quantities as number, location, and characteristics of sources, and even the sensor array configuration. This is related to such problems as ICA (independent component analysis), sensor fusion, and more standard array processing problems like finding direction of arrival. Our approach uses nonparametric estimates of mutual information to attempt to determine relationships between sensors, and frame the extraction of the original sources as an ICA-like problem.

Topic: **Clustering and efficient use of partially annotated information**
By: M. Szummer (MIT), T. Jaakkola (MIT)
IT-2, RCA-6
Efficient use of partially annotated/labeled data involves extracting structure from large unlabeled set and combining this information with limited labeled examples. A typical albeit unstated assumption in this context associates separable clusters in the unlabeled set with unique but unknown labels. When this assumption is valid, labeled examples are needed only to the extent that they can facilitate labeling of the clusters. We capture and formalize this intuition in a conditional probability model where soft clusters serve to regularize the labeling of the unlabeled examples. Clustering is achieved by defining a Markov diffusion process. The associated time scale of this process determines the effective size of the clusters and is chosen through a margin based criterion that guarantees unambiguous classification of examples. We relate the time scale to the mixing time of the Markov process and extend the basic idea by combining multiple time scales to maximize classification accuracy.

Topic: **Model geometry and estimation with incomplete data**
By: Adrian Corduneanu (MIT), Tommi Jaakkola (MIT)
IT-2, RCA-5
Dealing with limited or incomplete information is a recurring problem in large scale estimation tasks. Suitable constraints on the probability models permit a beneficial fusion of complete and incomplete information whereas under or overconstrained models may degrade as a result. We study the information geometry of probability models starting from simple overconstrained models and analyze when they can and cannot effectively exploit incomplete information in a classification context. We are also concurrently developing new estimation algorithms and criteria that are better suited for estimation tasks with predominantly incomplete data.

Topic: **Wireless Networking**
By: P. R. Kumar (UIUC)
IT-3, RCA-2 & 3
This research is aimed at studying the problems of scaling and capacity in ad hoc wireless networks. Such networks consist of a number of nodes which can communicate in a multi-hop fashion over the wireless medium.

Topic: A **Variational Approach to Array Processing in the Presence of Uncertainties**
By: Mujdat Cetin (MIT), Alan S. Willsky (MIT)
IT-2, RCA-1, with ties to RCA-4, 6
This research addresses sensor array processing problems by using an optimization-based framework. One such problem of interest is source localization in the presence of uncertainties in sensor locations. By exploiting an explicit model of the data collection geometry, sensor parameters, and statistically motivated constraints on the unknowns, the objective is to develop effective and robust variational schemes for information extraction.

Topic: **Inference on Graphs and Information Theory**
By: S. K. Mitter (MIT), A. Sahai (MIT)
IT-3, RCA-2 & 3
Publications: [4]
Sanjoy Mitter and his students are conducting research in Iterative Decoding and a Streaming View of Information Theory. Prof. Mitter has also developed a Variational View of Bayesian Estimation which is likely to be important for Inference on Graphs.

   Various portions of the research activities described above already either involve significant experimentation or will require it in the near future. As a result and also as is needed for RCA-7, we have initiated an examination of alternatives for data, experiments, transitions, and demos, including the development of some demo capabilities completely within our team and also exploiting our industrial partnerships to engage simultaneously in demonstration and transition activities.

References

[1] B. Frey, R. Patrascu, T. Jaakkola, and J. Moran. Sequentially fitting inclusive trees for inference in noisy-OR networks. In *Advances in Neural Information Processing Systems 13*, November 2000.

[2] T. Jaakkola. Tutorial on variational approximation methods. Inr *Advanced mean field methods: theory and practice*. MIT Press, September 2000.

[3] M. Meila and T. Jaakkola. Tractable Bayesian Learning of Tree Belief Networks. *Submitted*, December 2000.

[4] S. K. Mitter. Duality Between Estimation and Control. *Published in Festschrift for A. Bennoussan*, December 2000.

[5] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cascades of Gaussian scale mixtures and their use in modeling natural images with application to denoising. In *7th International Conference on Image Processing*, September 2000.

[6] M. J. Wainwright, E. Sudderth, and A. S. Willsky. Tree-based modeling and estimation of Gaussian processes on graphs with cycles. . In *Advances in Neural Information Processing Systems 13*, November 2000.